NPS55-88-007

NAVAL POSTGRADUATE SCHOOL
Monterey, California

AD-A200 471

VARIANCE REDUCTION USING NONLINEAR
CONTROL AND TRANSFORMATIONS

by
Peter A. W. Lewis
Richard L. Ressler
R. Kevin Wood

August 1988

88 11

Naval Postgraduate School
Monterey, California

Rear Admiral R. C. Austin   H. Shull
Superintendent   Provost
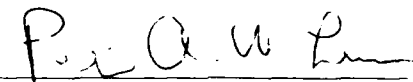
Reproduction of all or part of this report is authorized.

This report was prepared by:
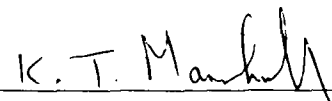
R. KEVIN WOOD   PETER A. W. LEWIS
Associate Professor of   Professor of Operations Research
Operations Research

Reviewed by: Released by:

PETER PURDUE   KNEALE T. MARSHALL
Professor and Chairman   Dean of Information and
Department of Operations Research   Policy Sciences

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b RESTRICTIVE MARKINGS		
2a SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited		
2b DECLASSIFICATION/DOWNGRADING SCHEDULE				
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NPS55-88-007		5 MONITORING ORGANIZATION REPORT NUMBER(S)		
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b OFFICE SYMBOL (If applicable) Code 55	7a NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000		7b ADDRESS (City, State, and ZIP Code)		
8a NAME OF FUNDING/SPONSORING ORGANIZATION Chief of Naval Research	8b OFFICE SYMBOL (If applicable) Code 512	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N0001487WR24023		
8c ADDRESS (City, State, and ZIP Code) Arlington, VA 22217		10 SOURCE OF FUNDING NUMBERS		

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR014-05-01	4114	

11 TITLE (Include Security Classification)

VARIANCE REDUCTION USING NONLINEAR CONTROLS AND TRANSFORMATIONS

12 PERSONAL AUTHOR(S)
Peter A.W. Lewis; Richard L. Ressler; R. Kevin Wood

13a TYPE OF REPORT Technical	13b TIME COVERED FROM ____ TO ____	14 DATE OF REPORT (Year, Month, Day) 88 August	15 PAGE COUNT 18

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Variance reduction; Control variables; ACE

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Nonlinear regression-adjusted control variables are investigated for improving variance reduction in statistical and system simulations. To this end, simple control variables are piecewise sectioned and then transformed using linear and nonlinear transformations. Optimal parameters of these transformations are selected using linear or nonlinear least-squares regression algorithms. As an example, piecewise power-transformed variables are used in the estimation of the mean for the two-variable Anderson-Darling goodness-of-fit statistic W_2^2. Substantial variance reduction over straightforward controls is obtained. These parametric transformations are compared against optimal, additive nonparametric transformations obtained by using the ACE algorithm and are shown, in comparison to the results from ACE, to be nearly optimal.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL Peter A. W. Lewis	22b TELEPHONE (Include Area Code) 408-646-2283	22c OFFICE SYMBOL 55Lw

DD FORM 1473, 84 MAR
83 APR edition may be used until exhausted
All other editions are obsolete

☆ U.S. Government Printing Office 1986—606-243

VARIANCE REDUCTION USING NONLINEAR CONTROLS AND TRANSFORMATIONS

Peter A. W. Lewis
Richard L. Ressler
R. Kevin Wood

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943.

August,15 1988

Abstract

Nonlinear regression-adjusted control variables are investigated for improving variance reduction in statistical and system simulations. To this end, simple control variables are piecewise sectioned and then transformed using linear and nonlinear transformations. Optimal parameters of these transformations are selected using linear or nonlinear least-squares regression algorithms. As an example, piecewise power-transformed variables are used in the estimation of the mean for the two-variable Anderson-Darling goodness-of-fit statistic W_2^2. Substantial variance reduction over straightforward controls is obtained. These parametric transformations are compared against optimal, additive nonparametric transformations obtained by using the ACE algorithm and are shown, in comparison to the results from ACE, to be nearly optimal.

1 PRELIMINARIES

This paper investigates the use of possibly nonlinear, regression-adjusted control variables for variance reduction in statistical and system simulation. Let $\underline{C}$ be a vector of control variables which are correlated with (related to or associated with) a statistic of interest, Y, and assume that $\underline{C}$ has a known mean vector $E[\underline{C}]$. The object is to more accurately estimate $E[Y]$ by deriving a controlled statistic Y'' which has less variance than Y. A standard method for doing this is via the linear, additive combination of Y and the components of $\underline{C}$,

$$Y' = Y - \underline{\beta}^T \left(\underline{C} - E[\underline{C}]\right). \tag{1}$$

1

The parameter vector $\underline{\beta}$ is a vector of unconstrained constants which are to be chosen so as to minimize the variance of Y'. Note that some components of $\underline{C}$ may be *known* power transformations of other components, so that polynomial control schemes are included in formulation (1). Explicit expressions for the components of $\underline{\beta}$ which minimize the variance of Y' can be found in terms of the second order moments of Y and $\underline{C}$, and with these parameters, Y' is an unbiased estimate of $E[Y]$.

In particular, consider the case of a single, additive, linear control $Y' = Y - \beta(C - E[C])$. Here β is chosen to minimize $var(Y')$. This variance is minimized when β is proportional to the correlation between C and Y: the greater the correlation, the greater the effectiveness of the control in obtaining variance reduction. Assuming $var(Y) = var(C)$, the result follows from:

$$var(Y') = var(Y) + \beta^2 var(C) - 2\beta cov(Y,C)$$
$$= var(Y)\left(1 + \beta^2 - 2\beta\rho(Y,C)\right).$$

Differentiating with respect to β and setting the resulting expression equal to zero yields the optimal value for β:

$$\beta = \rho(Y,C),$$

giving

$$\frac{var(Y')}{var(Y)} = 1 - \rho(Y,C)^2. \tag{2}$$

In particular,

$$100\left(\frac{var(Y) - var(Y')}{var(Y)}\right) = 100\left(1 - \frac{var(Y')}{var(Y)}\right) = 100\,\rho(Y,C)^2 \tag{3}$$

measures the percent variance reduction resulting from the control. Without the assumption of equal variances, we have

$$\beta = \frac{\sigma_Y}{\sigma_C}\,\rho(Y,C),$$

while (2) still holds. Thus if $\rho(Y,C)$, σ_Y, and σ_C are known, $\rho(Y,C)$ is a direct measure of the variance reduction which can be obtained with a single regression adjusted control. In fact comparing correlations is a method for choosing between proposed controls.

This paper generalizes (1) by letting

$$Y' = Y - C', \tag{4}$$

where C' is any mean-zero linear or nonlinear parametric function of the components of $\underline{C}$, i.e. $C' = f\left(\underline{C};\underline{\beta}\right) - E\left[f\left(\underline{C};\underline{\beta}\right)\right]$. For example, C' might involve additive or multiplicative combinations of unspecified power transformations of the components of the original control vector $\underline{C}$. Optimal or near-optimal values of the unknown parameters of these transformations, analogous to β in (1), are obtained by minimizing the variance of Y'. However, the results are not explicit functions of the joint and higher moments between Y and the set of control variables.

2

Now for this more general case of multiple, possibly nonlinear, control variables we obtain, using (4),

$$\frac{var(Y')}{var(Y)} = 1 + \frac{v\,r(C')}{v\,(Y)} - 2\frac{\sigma_{C'}}{\sigma_Y}\rho(Y,C')$$

$$= 1 + k^2 - 2k\rho(Y,C') \tag{5}$$

and

$$1 - \frac{var(Y')}{var(Y)} = 2k\rho(Y,C') - k^2, \tag{6}$$

where k is positive valued. While this last equation is simple in form, both $\rho(Y,C')$ and $k = \sigma_{C'}/\sigma_Y$ are functions of the parameters in C'. Thus it is not true that in order to maximize the variance reduction with respect to the parameters of the control function, one need only maximize the absolute value of the correlation between Y and C'.

When C' is a linear additive function of the components of $\underline{C}$ as in (1), $\rho(Y,C')$ is a quadratic function of the parameters $\underline{\beta}$ whose optimal values are a function of the correlation matrix of $(Y,\underline{C})$, i.e. the joint and higher moments between Y and the set of control variables. In fact, explicit expressions for the optimal values of $\underline{\beta}$ are known (Rubinstein and Marcus, 1985). As an example, for two independent linear controls with known correlations with Y, it follows from (5) that with the optimal values of $\underline{\beta}$,

$$\frac{var(Y')}{var(Y)} = 1 - \rho(Y,C_1)^2 - \rho(Y,C_2)^2. \tag{7}$$

Choosing control variables with maximum correlations with Y will, in this case, still maximize the reduction in variance.

In the general case (4), when the controls are not independent, and $\sigma_{C'}/\sigma_{Y'} \neq \rho(Y,C')$ in (5), $\rho(Y,C')$ does not yield an exact measure of variance reduction as does $\rho(Y,C)$ in (3). Additionally, the allowable range of parameters may be constrained in the function which generates C' out of the components of $\underline{C}$ by the requirement that $E[C']$ must be known, exactly or approximately, and must be finite.

2 THE ACE PROGRAM

The ACE (Alternating Conditional Expectation) program (Breiman and Freidman, 1985) provides a method for estimating the minimum variance obtainable by regressing a variable Y on an additive combination of arbitrary transformations of another set of variables such as the components of $\underline{C}$. As such it suggests that in a simulation context, more control (more variance reduction) can be obtained with transformations of the chosen control variables. ACE estimates transformation functions $\theta(Y)$ and $\psi_j(C_j)$ to minimize the fraction of the variance of Y not explained by regression (ϵ^2) defined as follows:

$$\epsilon^2(\theta, \psi_1, \ldots, \psi_p) = \frac{E\left[\left(\theta(Y) - \sum_{j=1}^p \psi_j(C_j)\right)^2\right]}{E[\theta^2(Y)]}. \tag{8}$$

3

The algorithm uses conditional expectations and alternates between improving the estimate of $\theta(Y)$ and improving the estimates of the $\psi_j(C_j)$s. The computational mechanics on finite data sets of continuous variables involve the use of data smooths to approximate the conditional expectations in order to repeatedly reduce the mean squared error until it is minimized.

The ACE procedure is nonparametric, with the transformations selected solely on the basis of the data sample. Minimal assumptions about the distribution of the sample or about allowable transformations enable ACE to produce an estimate of the minimum mean squared error between the transformed Y variable and the sum of the transformed components of $\underline{C}$. When $\underline{C}$ has only one component, this is equivalent to maximizing the correlation between a transformed Y and a transformed C.

Unfortunately, the transformations ACE selects cannot be used to develop control variables for variance reduction since the transformations are non-parametric and the true means of the transformed variables cannot be determined. However, one can use the minimum mean squared error from ACE to obtain an upper bound on the variance reduction that can be achieved between Y and C' in a parametric control function such as (4). Thus, ACE may be used to gauge the effectiveness of any control function using a fixed set of control variables. Since (4) does not allow for transformations of the dependent variable, ACE was intentionally limited during this study to using only linear transformations of Y.

3   THE SAMPLE ANALOG TO THE VARIANCE REDUCTION FORMULA

In practice, one has no theoretical information about the joint probability properties of Y and C'', or the joint probability properties of Y and the components of $\underline{C}$. Instead one has a simulation sample of size m of independent replications, $\{Y_i, C_i\} : i = 1, \ldots, m$, from which to estimate $E[Y]$. Regardless of whether the sample is large or small, i.e. is a pilot sample or all of the simulation data that will be available, one wants to minimize the sample variance of Y'. Minimizing the sample variance involves, after subtracting $\overline{Y}$ from both sides of (4), minimizing

$$\frac{\sum_{i=1}^{m} \left(Y_i' - \overline{Y} \right)^2}{m} \;=\; \frac{\sum_{i=1}^{m} \left(Y_i - \overline{Y} - C_i'^2 \right)}{m} \tag{9}$$

$$= \frac{\sum_{i=1}^{m} \left(Y_i - \overline{Y} \right)^2}{m} + \frac{\sum_{i=1}^{m} C_i'^2}{m} - \frac{2 \sum_{i=1}^{m} \left(Y_i - \overline{Y} \right) C_i'}{m} \tag{10}$$

The left-hand side of (9) is the quantity to be minimized as $E\left[\overline{Y}\right] = E\left[\overline{Y'}\right] = E[Y]$ since $E[C']$ is known to be zero. Thus either $\overline{Y}$ or $\overline{Y'}$ can be used in the estimate of the variance of Y'. Equation(9) shows that this estimate of the variance of Y' is equal to the residual sum of squares of the least squares regression of $Y_i - \overline{Y}$ on C'. Equation (10) involves, in its first term, the total sum of squares, which estimates the variance of Y; in its second term the sample variance of the zero mean C'; and in the last term the sample covariance of Y and C'.

4

Rearranging terms in (10), we have

$$\frac{\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2}{m} - \frac{\sum_{i=1}^{m} \left(Y_i' - \overline{Y}\right)^2}{m} = \frac{2\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right) C_i'}{m} - \frac{\sum_{i=1}^{m} C_i'^2}{m}$$

or

$$\frac{\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2 - \sum_{i=1}^{m} \left(Y_i' - \overline{Y}\right)^2}{\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2} = \frac{2\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right) C_i'}{\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2} - \frac{\sum_{i=1}^{m} C_i'^2}{\sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2}. \tag{11}$$

The left-hand side of (11) is the usual R^2 regression measure and equation (11) may be rewritten as

$$R^2 = 2\frac{S_{C'}}{S_Y} \, r(Y,C') - \frac{S_{C'}^2}{S_Y^2} = 2\hat{k} \, r(Y,C') - \hat{k}^2. \tag{12}$$

As the sample analog to (6), (12) indicates that maximizing R^2 through nonlinear least-squares regression is equivalent to maximizing sample variance reduction when the optimal parameters are unknown. Thus for $\underline{C}$ with multiple components, maximizing the effectiveness of C' can be accomplished through estimating the parameters of C' via multiple least-squares regression of $Y' - \overline{Y}$ on C'. A similar result relating optimal regression and optimal correlation can be found in the ACE article (Breiman and Friedman, 1985).

With linear controls, linear least-squares regression will provide a global minimum for the residual sum of squares, in turn maximizing the variance reduction for the sample. When the control function is nonlinear, nonlinear least-squares regression will not necessarily determine parameter values which globally minimize the residual sum of squares since the function could be nonconvex. With a control function $C' = f\left(\underline{C}; \underline{\beta}\right) - E\left[f\left(\underline{C}; \underline{\beta}\right)\right]$ that is nonconvex, there may be many suboptimal local minima. In this case the choice of initial values for the parameters $\underline{\beta}$ in the nonlinear regression may significantly affect the amount of variance reduction obtained. If one uses as starting values for $\underline{\beta}$ the special values which represent the linear case for the control, one should always do at least as well as the linear case regardless of nonconvexity.

One must be careful that while multiple regression may be computationally useful, the distribution theory behind multiple regression, which assumes fixed independent variables, does not apply. Consequently, confidence intervals on parameter estimates cannot be determined directly from the regression results.

4   APPROXIMATING OPTIMAL NONLINEAR TRANSFORMATIONS FOR NONLINEAR CONTROLS

Since ACE does not supply any parametric clue to the optimal transformations of the individual components of $\underline{C}$, approximations are needed for these transformations. A requirement for the approximations is that they contain the linear additive case (1) as a special set of parameter values, thus ensuring that one attains at least the known variance reduction for this case. The approximations studied here take two forms, piecewise linear controls, and standard statistical parametric transformations, used separately or conjointly on each component of C'.

4.1 Piecewise Linear Transformations of Controls

Statistics are often nonlinear functions of the random variables from which they are derived. Therefore one might expect some nonlinear controls to have a higher correlation with Y than linear controls. While not a measurable guarantee of improved variance reduction, (6) suggests that higher correlation indicates that a nonlinear control may be able to be a better "control" than the linear controls. Given an initial guess at a viable linear control, one type of nonlinear control can be formed by using indicator functions and "cutpoints" to form piecewise linear transformations of the control. Graphical analysis can be useful in selecting the initial cutpoint(s).

For example, a control variable C is split into two control variables about a cutpoint δ as follows:

$$C_1 = \begin{cases} C & \text{if } C \leq \delta \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad C_2 = \begin{cases} C & \text{if } C > \delta \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

By judicious choice of the variable cutpoint δ or perhaps multiple cutpoints, least-squares multiple regression can achieve a better fit without the use of additional control variables. As an example let X be distributed as an independent Uniform $(-.5, .5)$ variate. Let $Y = X^2 + \epsilon$ where ϵ is distributed as independent Normal $(0, 0.01)$. With 300 samples, using just X as a linear control as in (1), linear least-squares regression yielded an R^2 of 0.00. However using X to form two new controls as in (13), with $\delta = 0$, yielded an R^2 of .92 using linear least-squares regression. If $\delta \to \infty$ or $\delta \to -\infty$, the ordinary linear control is obtained. Of course, care must be taken in determining the form of the control function to ensure it has mean zero, i.e. $E[C_1]$ and $E[C_2]$ must be known. Note also that the regression is still linear if δ is *given*, but it is nonlinear otherwise. Finding an optimal δ then becomes, in general, a nonconvex, nonlinear, mathematical programming problem.

4.2 Transformations of Controls

Several standard transformations are used in statistics and data analysis and these can be applied to approximations for the optimal transformation of a control variable C. Power transformations of controls, in addition to piecewise transformations of controls, introduce nonlinearity into the controlled estimate of $E[Y]$ while containing the untransformed control as a special case. The power transformation used initially in this study is of the form $Z = (X^p - 1)/p$, for $p > -1$. This scaled power transformation has the property that as $p \to 0$ the limit is $\ln X$ and when $p = 1$ it gives a shifted version of the original variable.

This power transformation can have vastly different effects for $X > 1$ and $X < 1$. The curves in Figure 1 represent a sample of possible transformations. As one increases p, the change in the nature of the function on either side of $X = 1$ becomes more drastic. For large values of p, large values of X are given added weight while for small values of p, the small values of X are given the additional weight. Note that when $p = 1$, this is simply the linear transformation. Thus optimizing using this transformation assures variance reduction at least as good as in the linear case.
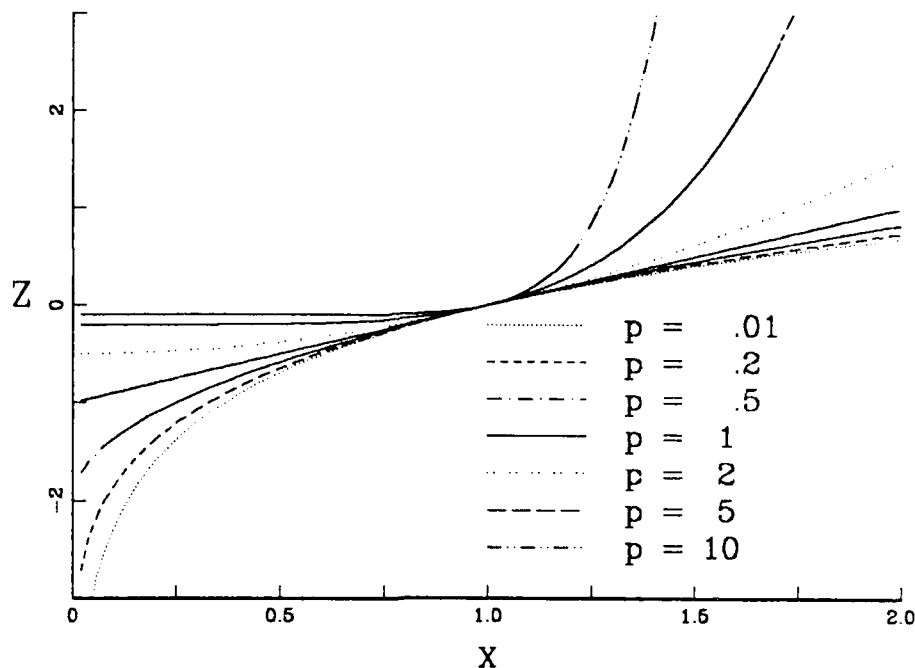
Figure 1: Examples of Power Transformations of a Variable X.

Using, for example, the single control variable C, the resulting nonlinear control function is

$$C' = \beta \left\{ \frac{C^p - 1}{p} - E\left[\frac{C^p - 1}{p} \right] \right\},$$

which has two parameters, p and β. Of course, combinations of piecewise transformations and power transformations are also possible, and it is this combination of nonlinear controls which is the main thrust of this paper. With this combination one hopes to come close to the maximum theoretical variance reduction which could be obtained.

5 AN EXAMPLE

Estimating the mean of the Anderson-Darling goodness-of-fit statistic, W_n^2, (Anderson and Darling, 1952) provides a good example of the benefits of piecewise internal controls and power transformations. The example is artificial since $E\left[W_n^2\right]$ is known to be one for all n, and the determination of the quantiles is the real problem. However, the example is useful as an illustration.

The statistic W_n^2 can be determined as a function of n independent unit exponential random variables E_j (Lewis and Orav, 1987). Note first that one can write W_n^2 as a function of order statistics from a unit exponential distribution as follows:

$$W_n^2 = -n - \left(n^{-1}\right) \sum_{i=1}^{n} \left[-(2i-1)\ln\left\{1 - e^{-E_{(i)}}\right\} + \{2(n-i) + 1\} E_{(i)} \right] \qquad (14)$$

where the $E_{(i)}$ are the order statistics from a unit exponential population. These order statistics can in turn be expressed in terms of the n independent unit exponential random variables E_j as

$$E_{(i)} = \sum_{j=1}^{i} \frac{E_j}{(n - j + 1)}. \qquad (15)$$

Together (14) and (15) give W_n^2 as a function of n independent exponential random variables. The independence of these random variables makes them ideal for controlling W_n^2. The case $n = 2$ is presented here, for which (7) holds with $C_1 = E_1$ and $C_2 = E_2$.

As mentioned before, graphical methods can sometimes be useful in determining types of controls or aspects of controls. Two useful plots of W_2^2 are presented here. Figure 2 is a surface plot of W_2^2 over a small region of the E_1, E_2 plane where the majority of values occur. This is not a density plot, but a representation of the functional relationship between the two independent exponentials and the W_2^2 values each pair generates. Subsequent surface plots of the control functions likewise do not portray density: just the surface generated by the control function. As an indicator of the density of points on the W_2^2 surface, Figure 3 is a sample bivariate histogram of 1000 independent pairs of unit exponentials. While one could plot an actual bivariate exponential density, the discrete nature of the histogram allows easier visual comparisons of density. Together, Figures 2 and 3 indicate why nonlinear controls may prove useful for controlling W_2^2. Clearly the relationship between W_2^2, E_1 and E_2 is highly nonlinear suggesting the use of nonlinear controls. Figure 3 supports one's intuition that the majority of pairs of the bivariate exponential are close to the origin. Suspecting this one may be tempted to use a linear control to just approximate the surface in this region. However Figure 3 also shows a significant number of pairs throughout the plane. Thus in order for C'' to be an effective control, the entire surface should be approximated by the control. This would require a nonlinear control and nonlinear regression.
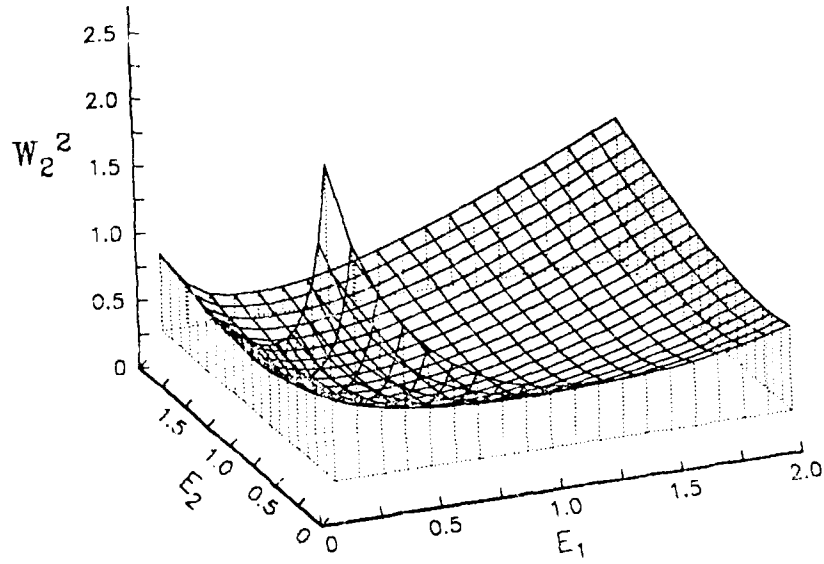
8

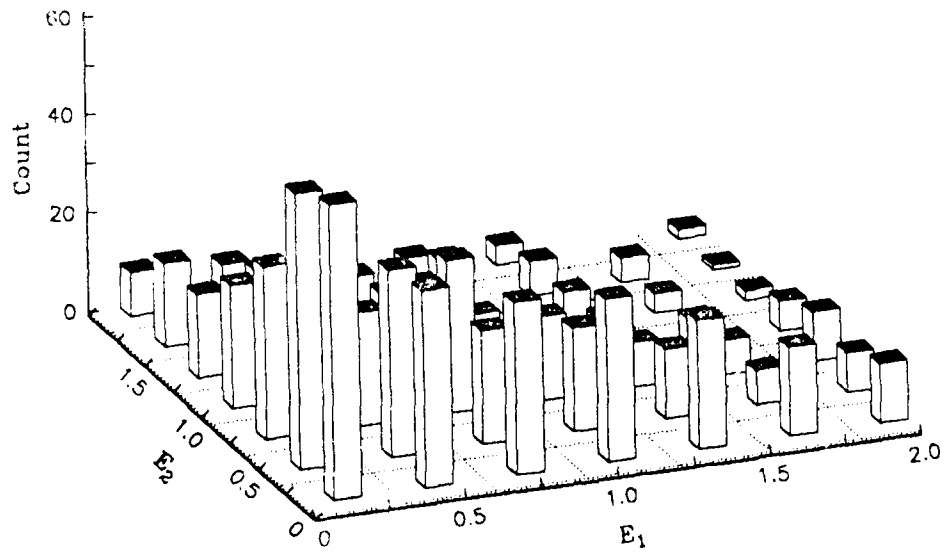Figure 2: The Nonlinear Surface of W_2^2 as a Function of Two Variables E_1 and E_2.

Figure 3: 1000 Samples of E_1, E_2 Pairs

9

Six different linear and nonlinear control functions for estimating the mean of W_2^2 were evaluated using a single sample of 500 pairs of unit exponentials and their associated W_2^2 values. The experimental, APL-based GRAFSTAT, from IBM Research, was used for all of the computing. The following six control functions were compared:

$$C' = \beta_1 (E_1 - 1) + \beta_2 (E_2 - 1);$$ (16)

$$C' = \beta_1 (E_1 - 1) + \beta_2 (E_2 - 1) + \beta_3 (E_1^2 - 2) + \beta_4 (E_2^2 - 2);$$ (17)

$$C' = \sum_{j=1}^{2} \beta_j \left[\frac{E_j^{P_j} - 1}{P_j} - E\left[\frac{E_j^{P_j} - 1}{P_j} \right] \right];$$ (18)

$$C' = \beta_1 (E_1 - 1) + \beta_2 (E_2 - 1) + \beta_3 \left(\frac{E_1^{P_1} - 1}{P_1} - E\left[\frac{E_1^{P_1} - 1}{P_1} \right] \right) + \beta_4 \left(\frac{E_2^{P_2} - 1}{P_2} - E\left[\frac{E_2^{P_2} - 1}{P_2} \right] \right);$$ (19)

$$C' = \sum_{j=1}^{2} \sum_{k=1}^{2} \beta_{jk} \left[\frac{E_{jk}^{P_{jk}} - 1}{P_{jk}} - E\left[\frac{E_{jk}^{P_{jk}} - 1}{P_{jk}} \right] \right],$$ (20)

where

$$E_{j1} = \begin{cases} E_j & E_j \leq \delta_j \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad E_{j2} = \begin{cases} 0 & E_j \leq \delta_j \\ E_j & \text{otherwise} \end{cases} \quad j = 1,2;$$

and

$$C' = \sum_{j=1}^{2} \sum_{k=1}^{3} \beta_{jk} \left[\frac{E_{jk}^{P_{jk}} - 1}{P_{jk}} - E\left[\frac{E_{jk}^{P_{jk}} - 1}{P_{jk}} \right] \right],$$ (21)

where

$$E_{j1} = \begin{cases} E_j & E_j \leq \delta_{j1} \\ 0 & \text{otherwise,} \end{cases} \quad E_{j2} = \begin{cases} E_j & \delta_{j1} < E_j \leq \delta_{j2} \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$E_{j3} = \begin{cases} E_j & E_j > \delta_{j2} \\ 0 & \text{otherwise} \end{cases} \quad j = 1,2.$$

As expected, the simplest controls, (16) and (17), with straightforward control functions, were usually less effective than the nonlinear controls. The controls give, by (16) and (17) are referred to as the "standard" controls because their unknown parameters can be computed using linear least-squares regression. Since the necessary expected values of the controls just involved the first two moments of the exponentials, they were determined analytically and not estimated. The remaining parameters for controls (16) and (17), respectively β_1 and β_2, and $\beta_1, \beta_2, \beta_3$ and β_4 were computed using linear least-squares regression. Since control (16) is a linear function of E_1 and E_2 and W_2^2 is a very nonlinear function of E_1 and E_2, this control, not surprisingly, achieved an R^2 of only .2265. This poor performance could also be predicted by using the sample estimates for $\rho(W_2^2, E_1)$ and $\rho(W_2^2, E_2)$ in (7). If the estimates were the

10

true correlations, the optimal β's would only yield a 22.66 percent variance reduction. The parabolic shape of (17), as shown in Figure 4, enabled the control function to achieve an R^2 of .5627. While better, it is far from optimal. Note that on the graphs of the controls the predicted values of the controls are centered about zero, the mean of C'.
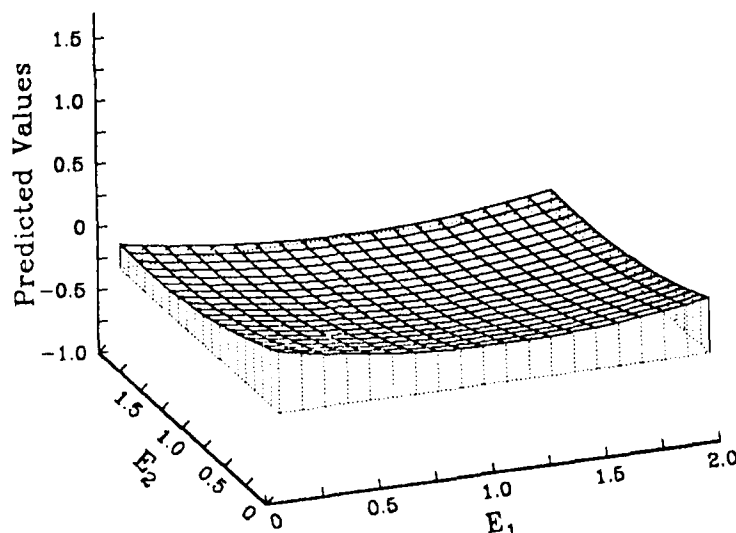

Figure 4: Surface generated by the parabolic linear control given in (17). The control is linear since the powers are fixed.

For control (19) only the linear terms' expected values could be calculated analytically. The other two expected values were functions of the unknown parameters and had to be recalculated based on the current parameters during the optimization. For controls (18), (20), and (21) none of the expected values could be determined analytically so all were calculated during the optimization. All of the parameters for the nonlinear controls were estimated via the nonlinear regression segment of GRAFSTAT. For nonlinear regression, GRAFSTAT uses a form of the Marquadt algorithm (Marquadt, 1963) which allows bounds to be placed on the parameters. Lower bounds of -.99 were necessary on the power parameters, p_{jk}, since the expected values of the exponentials (involving the gamma function) are not defined for $p_{jk} \leq -1$. A reasonable upper bound on each p_{jk} was found useful in speeding convergence.

As the control functions became more nonlinear, their effectiveness usually increased. Allowing the powers to float in control (18) versus being fixed in control (16) gave a slight improvement: the R^2 went from .2265 up to .4640. This was not as good however as the "standard" control (17) with two linear terms and two quadratic terms which achieved an R^2 of .5627. Adding the two linear terms to control (18) resulted in control (19). Now allowing the powers to float in control (19), versus being fixed in control (17), enabled the surface to fit more closely and thus the R^2 for (19) was .7422. This definite improvement over the "standard" controls can be seen in Figure 5.
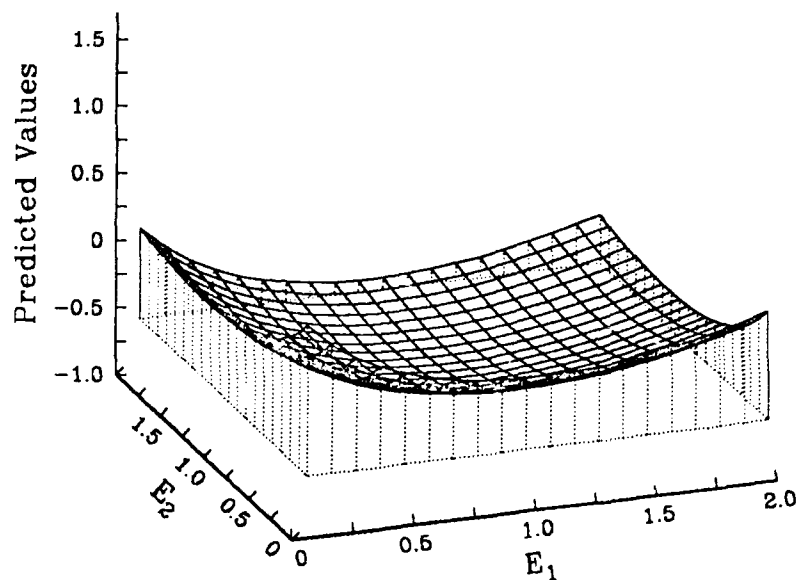
11

Figure 5: Surface generated by "non-standard" control with linear and nonlinear terms given in (19).

Originally, the cutpoints for controls (20) and (21) were parameters to be optimized. Unfortunately, this made the optimization unstable and the results unreliable. Thus, the cutpoints were fixed at selected quantiles and not included as parameters in the nonlinear regression. Selection of a good cutpoint was done by examining the results of a short sequence of regressions. For control (20) a cutpoint at the .5 quantile was the most effective one found for this sample. Comparing Figure 6 to Figure 5 shows the impact of adding nonlinearity by the use of the cutpoint. The R^2 for control (20) was .8216. The results of using the estimated parameters for (20) on three independent samples of 1000, Table 1, indicate that even though the regression-estimated parameters are biased for the original sample, (20) is still effective in controlling other samples.

	Sample 1	Sample 2	Sample 3
$\bar{Y}$	.9882	1.0022	1.0219
$s_{\bar{Y}}$	.0261	.0262	.0282
$\bar{Y}'$	.9972	1.0095	1.0238
$s_{\bar{Y}'}$	.0110	.0124	.0129
R^2	.8239	.7759	.7905

Table 1: Effect of the nonlinear, single-cutpoint control given in (20) on three independent samples other than the regression sample.
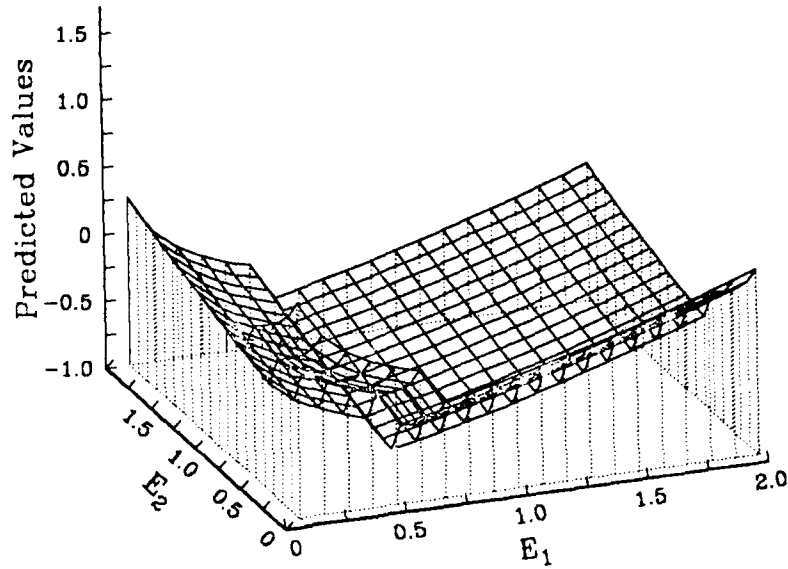
12

Figure 6: Surface generated by the nonlinear, single-cutpoint control given in (20).

As the number of cutpoints increases to two for control (21), one gets a more effective control at the cost of increased computational complexity. The computational complexity increases because the additional cutpoint creates more parameters and because the computation of expected values becomes more expensive. As before, the cutpoints were fixed at selected quantile values. Which values to select was a matter of performing a series of regressions on a grid of values. Figure 7 shows that some pairs of cutpoints were better than others. Figure 8 shows that the best cutpoints for this sample on the grid examined, the .30 and .65 quantiles, yield a control which is an excellent approximation to the $W_2'^2$ surface. The regression with these cutpoints on the original sample yielded an R^2 of .8372.

This last control, (21), was then tested on independent samples and the R^2 was compared to results from ACE. Table 2 indicates the results for three samples of 1000 W_2^2 values. Again the R^2 values are almost as good as the original sample, and the control is effective in all three cases. ACE was given the data generated by using the cutpoints on the original sample as the independent variables. The R^2 value derived by ACE was .8560 showing that control (21) is nearly optimal for the control variables used.
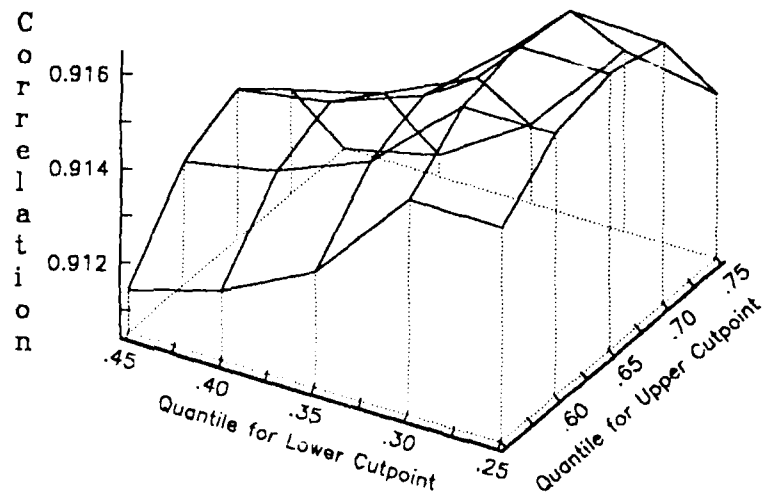
13

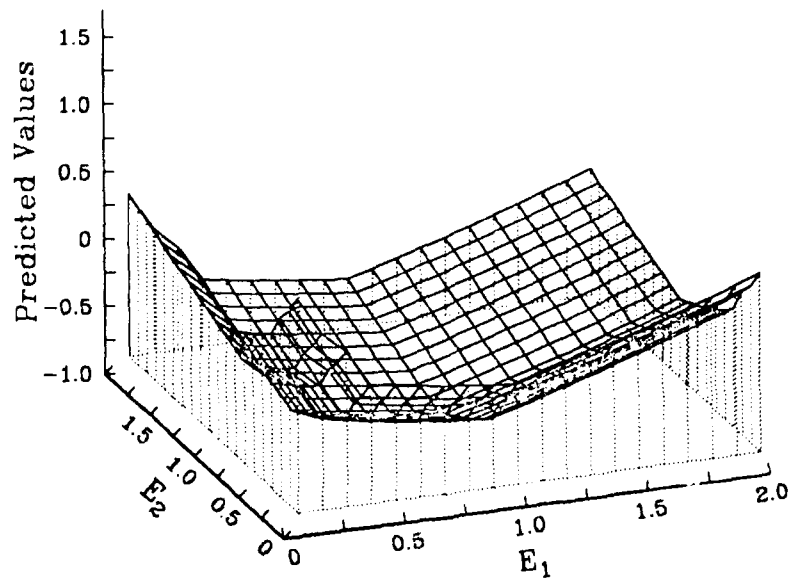Figure 7: Effects of Changing the Cutpoints on Correlation

Figure 8: Surface generated by nonlinear, double-cutpoint control given in (21).

14

	Sample 1	Sample 2	Sample 3
$\bar{Y}$	.9534	1.0230	.9842
$s_{\bar{Y}}$	.0230	.0265	.0255
$\bar{Y}'$	.9997	1.0157	1.0001
$s_{\bar{Y}'}$	.0094	.0121	.0110
R^2	.8350	.7925	.8153

Table 2: Effect of the nonlinear, double-cutpoint control given in (21) on three independent sample other than the sample used for regression.

6 SUMMARY AND CONCLUSIONS

This study demonstrates the potential effectiveness of nonlinear regression-adjusted controls in reducing variance in simulations. Various piecewise linear and power transformations were shown to be useful in developing control functions. There are many questions yet to be answered though. Some areas of investigation are listed.

1. Finding controls for the variance, percentiles and quantiles of W_2^2. Once a suitable control function is developed, can it be used with different parameters for other aspects of the data?

2. Finding controls for W_n^2 for $n > 2$. As the dimensionality increases, one may not need every independent variable in the control function to get effective control. Measures of influence or leverage could possibly be used to reduce the size of the control function.

3. Using other transformations such as

   (a) $Z = \left(e^{\gamma X} - 1 \right) / \gamma$.

   (b) $Z = \left((X^p - 1) \left(e^{\gamma X} - 1 \right) \right) / p\gamma$, or

   (c) $Z = \left(e^{\gamma(X^p - 1)/p} - 1 \right) / \gamma$.

   These transformations represent a broad spectrum of transformations on a variable as can be seen in Figures 9, 10 and 11. Note also that transformation (3a) and transformation (3c) contain the linear case as a special set of parameter values. The first transformation, (3a), is a positive weighting of all values, with large values weighted more than small values. By varying the γ parameter, one can scale the effects of the weights from very large for large γ to very slight for small γ. The second transformation, (3b), applies small negative weights for values less than 1. For values larger than 1 it allows for a wide range of positive weighting schemes as in transformation (3a). The third transformation, (3c), is similar to the straightforward power transformation, (Figure 1), but with more parameters. Thus it allows for more flexibility and increased curvature for smaller values of the parameters. The difficult part with these transformations, as usual, is computing the necessary expected values.
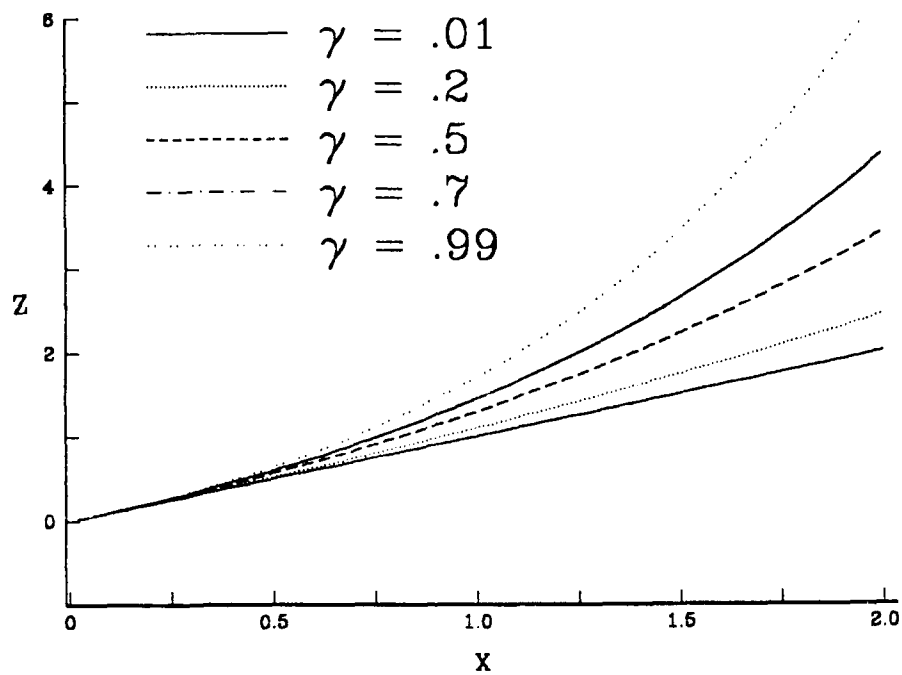
15

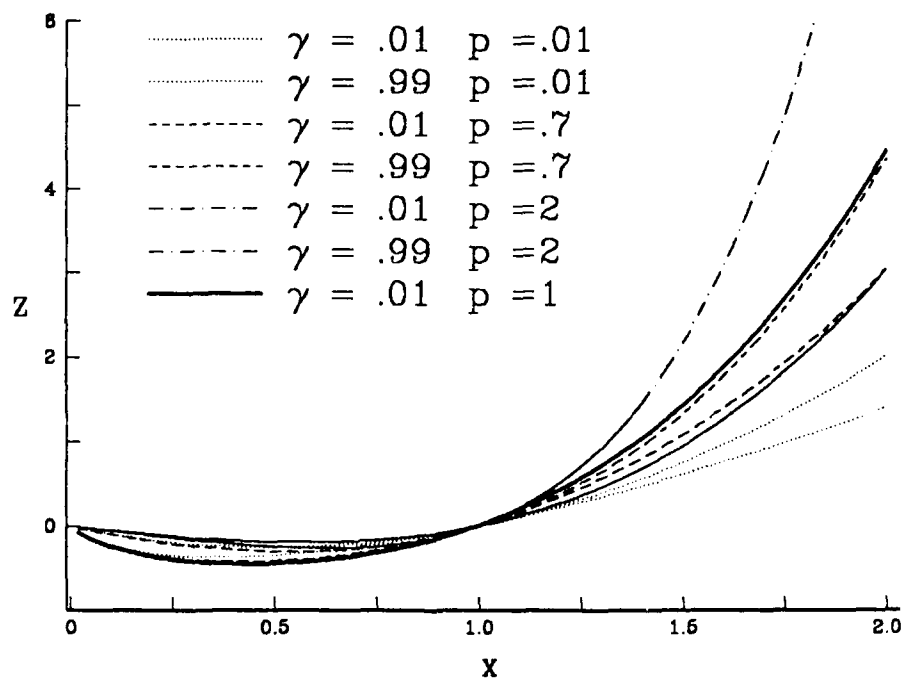Figure 9: Transformation (3a) applied to a Variable X.

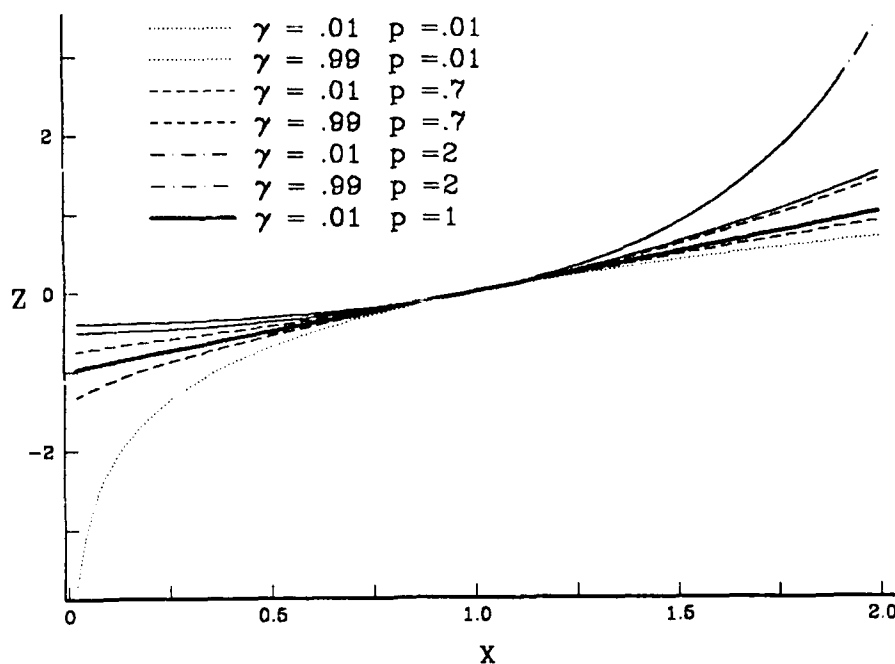Figure 10: Transformation (3b) applied to a Variable X.

16

Figure 11: Transformation (3c) applied to a Variable X.

4. Using similar controls for gamma family statistics such as those encountered in queuing problems. Preliminary results with internal control of a regenerative simulation estimate of the waiting time of the nth customer in an M/M/1 queue (Iglehart and Lewis, 1979) indicate that allowing non-integer powers in the control function can substantially increase the effectiveness of the control. On a sample of 20,000 busy periods, with $\rho = .5$ and $\mu = 1$, a linear control function obtained an R^2 of .59 while a nonlinear control function obtained an R^2 of .69.

5. Investigating problems with estimating the variance of the variance-reduced estimate of $E[Y]$. This is a very difficult problem for which sectioning or bootstrapping may be needed.

ACKNOWLEDGEMENTS

17

REFERENCES

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic process, *Annals of Mathematical Statistics*, **23**, 193-212.

Breiman, L. and Freidman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association*, **80**, 580-619.

Iglehart, D and Lewis, P.A.W. (1979). Regenerative Simulation with Internal Controls, *Journal of the Association for Computing Machinery*, **26**, No. 2, 271-282.

Lewis, P.A.W. and Orav, E.J. (1987). *Simulation Methodology for Statisticians, Engineers, and Operations Analysts*, to be published.

Lewis, P.A.W., Ressler, R.L. and Wood, R.K. (1987). Smarter Control Variables: Regression-Adjusted Linear and Nonlinear Controls, *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, H. Grant, W.David Kelton (eds.), 342-346.

Marquadt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal*, **11**, 431-441.

Rubinstein, R.Y. and Marcus, R. (1985). Efficiency of multivariate control variates in Monte Carlo Simulation, *Operations Research*, **33**, 661-667.

DISTRIBUTION LIST

	No. of Copies
Professor Peter Purdue, (Code 55Pd) Naval Postgraduate School Monterey, CA 93943-5000	1
Professor P.A.W. Lewis, (Code 55Lw) Naval Postgraduate School Monterey, CA 93943-5000	10
Professor Kevin Wood,(Code 55Wd) Naval Postgraduate School Monterey, CA 93943-5000	10
Mr. Richard Ressler, (Code 55), Naval Postgraduate School Monterey, CA 93943-5000	10
Arthur P. Hunter, Jr. Professor and Chairman Department of Industrial Engineering and Management Sciences Northwestern University Evanston, IL 60201-9990	1
Koh Peng Kong OA Branch, DSO Ministry of Defense Blk 29 Middlesex Road Singapore 1024	1
Commanding Officer ATTN: Mr. Murray Rowe Navy Personnel Research and Development Center San Diego, CA 92152	1
Commanding Officer ATTN: Barry Siegel Navy Personnel Research and Development Center San Diego, CA 92152	3

1

	No. of Copies
Mr. R. Morton %DSAI 350 Fortune Terrance Rockville, MD 20854-2995	4
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Office of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analyses 4401 Ford Avenue Alexandria, VA 22302-0268	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Operations Research Center, Rm E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J.F Shapiro Cambridge, MA 02139	1
Institute for Defense Analysis 1800 North Beauregard Alexandria, VA 22311	1